WORKING PAPER
ALFRED P. SLOAN SCHOOL OF MANAGEMENT

Experimentation in Learning Organizations: A Management Flight Simulator Approach

by

Bent **BAKKEN**, Janet **GOULD**, and Daniel **KIM**

WP-3274-91-BPS

MASSACHUSETTS
INSTITUTE OF TECHNOLOGY
50 MEMORIAL DRIVE
CAMBRIDGE, MASSACHUSETTS 02139

D-4203

Experimentation in Learning Organizations: A Management Flight Simulator Approach

by

Bent **BAKKEN**, Janet **GOULD**, and Daniel **KIM**

WP-3274-91-BPS

Experimentation in Learning Organizations: A Management Flight Simulator Approach.

Bent BAKKEN, Janet GOULD, and Daniel KIM

MIT Sloan School of Management, E40-294, Cambridge, MA 02139, USA

Abstract: Managers' real life experiences will need to be supported by new learning tools, as external environments and internal dynamics of organizations become more complex. This need for simulated managerial experience has come from a trend towards fewer hierarchical levels in organizations. Managerial competency that was once achieved by progressing through the many layers needs to be obtained through other means. This article shows how management flight simulators can enhance learning by allowing managers to compress time and space, experiment with various strategies, and learn from simulated deployments by reflecting on the outcomes. We focus on three aspects of the use of such simulators 1) the design of a generic learning *process*—learning laboratories—to address service quality management issues in a variety of industrial settings; 2) how dimensions of this learning can be operationalized and *measured*; and 3) an experimental study that shows the *transfer* aspect of learning and how different learning strategies affect the way in which participants *transfer* learning across domains.

Keywords: Learning organizations, transfer, gaming, system dynamics, behavioral decision making, simulation

1. Introduction

Organizational decision making is highly complex and managerial choices are far from trivial. The fact that people have a hard time learning from real-life experiences compounds the problem, especially when the decisions and their consequences are separated in time and space. In the US, for example, one of New England's largest banks provides an illustration of how difficult learning from experience can be. The bank's aggressive pursuit of ship-owning clients in the seventies led to huge losses in shipping portfolios in the early eighties. Massive losses led the bank to liquidate its entire shipping portfolio. At the same time, however, they aggressively expanded into risky real estate developments. In the late eighties, their real estate ventures resulted in quarterly losses of up to one billion dollars. These losses eventually caused the bank's ultimate demise in 1991. Although shipping and real estate markets share many structural features and are unstable for similar reasons, this bank was apparently unable to learn from their mistakes in shipping and repeated them in real estate (Bakken, 1990).

This example not only highlights how difficult it is to learn causal relations in one decision environment, but also shows that decision makers do not easily draw appropriate lessons from failures. Certainly, people must simplify their decision environments into manageable chunks, lest decision making be impossible (Simon, 1956; 1978). Yet, these simplifications do not work very well—people are poor intuitive judges and tend to violate

almost all rules of rationality and consistency (Kahnemann and Tversky, 1974). Decisions are prone to error even in quite simple contexts (Brehmer, 1989; Sterman, 1989). People often fail to grasp the power of exponential growth and pay too little attention to supply line information.

In experiments as well as in real organizational settings, decision makers have access to feedback about the appropriateness of actions (Einhorn and Hogarth, 1978). As our bank example shows, however, the feedback interpretation in slowly evolving environments is far from straightforward. Many causal inferences are possible, and only an experimental approach can sort out the many competing explanations. Unfortunately, organizational life does not lend itself readily to experimental testing, leaving organizations and their members to construct meanings out of what they experience. Many organizational researchers have come to the same conclusion—organizational environments and meanings are not given, but constructed. In this construction process, intra-organizational defense mechanisms come to play, and as a consequence, people bias their choice of information sources (Argyris and Schön, 1978; Weick, 1977) thus preventing learning from taking place.

The use of feedback can only be effective if cause and effect are closely related in time and space, but real decision environments lack this closeness between decisions and meaningful feedback. In real estate markets, for example, it takes three to four years from the time a decision to build an office building is made until people can move into the building. In the meantime, markets may have changed significantly. Although the Total Quality movement has successfully capitalized on feedback information, by making cause and effect closer in time and space, (Kim, 1989) there are many instances where shortening feedback cycle times is difficult. As the real estate case shows, there are important physical constraints on how fast feedback can become available.

Current changes in corporate structuring, such as de-layering of management levels, suggest that learning from real world experience will become increasingly more problematic. A beneficial outcome of such recent trends toward flatter organizational structures is that feedback delays are shortened, making organizations more responsive. Fewer hierarchical levels, however, mean that there are fewer "training steps" on the corporate ladder. Instead of spending decades in the same organization at various levels of responsibility, the "lean and mean" organization catapults the typical manager into decision making authority much sooner and without the extensive experience typically associated with senior decision makers.

Thus, problems inherent in learning from unguided organizational experience are compounded by ever scarcer time available for learning at a time when increasing

interdependencies make learning more important. The problem is further heightened in cases where real life feedback cycles are much longer than the decision making cycle. In such cases, computerized Management Flight Simulators can provide *virtual* worlds in which assumptions, relationships, and outcomes can be tested, thereby shortening the feedback cycle time in situations where delays are inherently long.

A Management Flight Simulator is a learning tool that allows managers to compress time and space, experiment with various strategies, and learn from making rounds of simulated decisions. Embedded in specially designed learning environments called learning labs, these simulators can be designed to provide organizational "practice fields" (Senge, 1990) where managers can experiment and learn in environments that allow failure and reflection.

2. Philosophy of the Learning Organization and the Role of Flight Simulators

Increasingly, attention is being focused on finding ways to make corporations more responsive to customers and enable them to provide higher quality products (de Geus, 1988; Stata, 1989). A "learning organization" is characterized by its attention to enhancing thinking processes that lay behind decision making. Furthermore, learning is encouraged through experimentation and testing in virtual worlds (Senge, 1990). Although learning organizations have attributes along many dimensions, the interesting point for our purposes is that (a) they have an exploratory attitude and (b) current solutions and processes are open to questioning (Argyris and Schön, 1978). Organizations need to experiment in both production and management processes while balancing the need to be as close to the implementation world as possible.

As mentioned earlier, Management Flight Simulators can be used in situations where real life experimentation is unfeasible because of cost considerations, time involvement, or both. They can be regarded as a framing tool for dynamic issues, often referred to as transitional objects (Papert, 1981). Moreover, thinking can be more structured and discussions more productive since discussions can be focused around a computer model that helps de-personalize assumptions and makes them less threatening.

For the same reason that we find Flight Simulators appealing—they are low cost, allow compression of time and space, and are conducive to reflection—there is a long tradition in the arts, sciences and professions of using similar "virtual worlds" (Schön, 1983). In fact, in technical professions, while "virtual worlds" have always existed, recent advances in simulations and other tools have made them ever more powerful. With the exception of spreadsheet analysis on personal computers, however, little has been made of

virtual worlds for improving managerial practice. This lack of the widespread use of management simulators may be attributable to two flaws—most are either too simplistic to feel "real" or too complex to learn from. Simulators based on system dynamics overcome these polar extremes by providing a framework with which participants can make sense of complexity and by producing realistic and challenging dynamics (Vennix, 1990). Management Flight Simulators can help in sorting out competing explanations by allowing participants to conduct experiments and learn from them.

In this article, we provide an in-depth look into the organizational processes by which simulator learning can be achieved. In doing so, we emphasize the need to provide a non-threatening environment. The first section shows our underlying theory for designing and running simulator-based workshops. A generic service-quality, management flight simulator is used as a tool for questioning assumptions that impede learning in the real organizational setting. In the next section, we define the dimensions of simulator-based learning and how to measure the learning. Several types of instruments are described. In the final section, we focus on the transfer dimension of such learning.

3. Designing a Reflective Learning Environment

In order for Management Flight Simulators to be effective learning tools, they must be designed into an environment that is conducive to learning. Without such an environment, the simulators become more of a management video game, and the goal becomes one of beating the highest score. The goal of a learning laboratory is to provide an environment that will help enrich managers' mental models using tools such as the management simulators. Learning laboratories help managers leverage their domain-rich knowledge by allowing them to play through simulated years, reflect on their actions, modify their mental models, then repeat the process. By compressing time and space, flight simulators can accelerate learning by enabling them to conduct many such cycles of action and reflection.

The impetus for designing such a learning laboratory originally came out of a study conducted at a major property and casualty insurance company. (Geraldine Prusko and Robert Bergin of Hanover Insurance were largely responsible for the overall design of the initial learning laboratories at this insurance company.) The goal of the design team was to create an environment in which managers could step out of day to day demands to reflect on their decision-making (for a complete description, refer to Senge and Sterman in this journal). In addition, managers would also develop a common language, learn new tools for thinking systemically, discuss operational objectives and strategies in an open forum, test operating assumptions, experiment with new ideas about managing a claims office, and have fun while doing it.

3.1. A Generic Learning Laboratory Design

To leverage the benefits derived from a case-specific lab, such as the insurance claims learning lab, both the computer game and the workshop design were modified to be more generally applicable. What follows is a description of a generic learning laboratory on Service Quality Management that has been adapted from the original claims learning laboratory design. This generic design can be used as the basis for creating other learning laboratories in domains outside the insurance industry. In addition to design issues, process points are also included to help the reader gain a better feel for the learning laboratory experience.

The overall design of the service quality management learning laboratory has four distinct stages: context setting, conceptualization, and experimentation and reflection. Each of the stages will be described.

Context Setting

Workshop leaders in the learning laboratory are positioned from the outset as "enablers", not authority figures. The participants are encouraged to challenge the assumptions of the model that underlies the simulators used in the learning lab. This openness to challenge and to test is critical for establishing a common understanding between workshop leaders and participants. The learning laboratory is positioned not as an answer generator but as a useful vehicle for illuminating and communicating issues of importance. Participants are also encouraged to share any reservations or concerns they may have about the laboratory with the rest of the group. These steps emphasize the experimental aspect of the laboratory and encourage participants to challenge their own operating assumptions.

In small groups, participants are asked to identify a specific industry setting where service quality management is important. The service can be an internal service or one with external customers. There are five key variables that the participants must fill in with specific factors that are relevant for their chosen industry setting. They must identify the Personnel and Customers, the Service, and the Direct costs and Cost of Poor Quality. For example, in a banking industry context (see Figure 1), personnel would be loan officers, the customers would be individual or business borrowers, the service "product" would be various types of loans, the direct costs would be personnel expenses including overhead, and the cost of poor quality would result in expenses resulting from loan delinquencies.

context / variables	insurance claims office	bank loan office	engineering department	internal training & development
Personnel	adjusters	loan officers	engineers	sales training staff
Customer	claimants	individuals, companies, developers	internal producers of designed components	world sales management group
Service	adjusting losses	loans	engineer traction control systems	comprehensive dynamic training
Direct Costs	employee salaries & overhead	employee salaries, overhead, transaction	salaries, testing time, prototype materials, testing equipment	sales training budget, participant travel expense
Cost of Poor Quality	high settlement costs, including litigation expenses	nonperforming loans, defaults	redo job, missed schedules, customer satisfaction, warranty costs, cost overruns	wasted time of attendees, poor sales service to external customers, lost sales

Figure 1. Examples of various contexts used in a generic learning lab on Service Quality Management

Working in small groups, the participants are asked to brainstorm and come up with the setting and variables described above. The overall purpose of this exercise is to have everyone think in terms of specific issues that they can personally relate to and establish a common understanding of the industry setting the learning laboratory is intended to address.

Conceptualizing the Issues

The groups of participants are asked to define what Quality is within the context they have chosen and to identify the key factors that determine quality. Causal loop diagramming tools are introduced as a way of representing the interconnectedness of a system. With each round of conceptualizing, small portions of the causal structure embodied in the simulator model are presented as a way of connecting the tool to the issues at hand.

Participants then focus on a particular issue (e.g., one of the decision variables in the flight simulator) 1) determine the key factors that affect that variable; 2) sketch patterns of behavior; 3) provide structural explanations (using causal loop diagrams); and 4) identify possible intervention points. One person in each group is responsible for presenting the work to the entire group. This process is then repeated, addressing a different issue.

Conceptualizing these variables, the participants draw from their domain-specific experience and explicate their mental models. At the same time, the participants are replicating part of the model-building process, making it easier to accept and identify with the pre-developed model as it becomes less and less of a black box. Each presenter "tells a story" with a causal loop diagram and illustrates with a real world example, if possible. This process helps participants develop the ability to articulate causal structures to other people. The overall objective in this stage is to have the group cover all the major issues contained in the management flight simulator and have a chance to challenge and test the inter-relationships that different people within the group may propose.

Experimentation and Reflection: "Flying" the Flight Simulator

Participants are then grouped in teams of two and are instructed to pursue a single-minded goal where they are accountable for meeting one particular goal (e.g., maintaining headcount). In these planned scenarios, they are told in advance that there will be a one-time 20% step increase in workload. For each simulated month, three decisions must be made—hiring, desired production, and quality goal. The hiring decision is for adding or reducing the number of personnel. Setting desired production at 1.00 translates into asking the employees to have the outflow of work equal the inflow of new work received for each month. A productivity goal of less than 1.00 means the backlog of customers is growing since we are servicing less than the number coming in. The quality goal can be entered directly as a decision point where a quality level of one is simply a starting reference point against which one can gauge improvements or deterioration.

These planned scenarios allow the group to gain experience slowly by trying a very focused strategy through which they can get a feel for the dynamics of the simulator. Optimally, the debriefing structure encourages more reflection in every phase of the process—strategizing, managing within the simulator, and debriefing the outcomes. A more important underlying purpose is to begin addressing particular organizational issues through appropriate choices of planned scenarios—continually connecting refinements in mental models back to their work domain.

In later sessions, the participants manage their simulated offices where they have no *a priori* knowledge of how the stream of new customers is going to change. In both the planned scenarios and the free plays, each team is asked to do the following for every trial 1) Plan a strategy and commit to it on paper; 2) predict the consequences of executing the strategy by sketching in behavior over time of some key variables; 3) play the game; and 4) debrief the game results and explain them to the rest of the group.

The discipline of planning a strategy and sketching out anticipated behavior in advance is important because it forces them to get in the habit of making their mental models explicit. The written records also provide a basis for comparing the actual outcome of each trial. By making such comparisons, gaps in thinking or failures to follow through an intended strategy are highlighted.

3.2. Unsurfacing Hidden Assumptions

If we view learning as a process where an *action--> result--> reflection--> learning* leads back to further *action*, Flight Simulators can facilitate learning by shortening the delay between *action-->result*. The simulator also demands structural explanations of the *action-->result* link that will force participants to search for a better understanding of the underlying forces that produce a given set of outcomes. The design of the learning laboratory also increases *reflection* and enhances *learning* out of which better decisions can arise.

How the learning laboratory can be used to engage the participants in what Argyris and Schön (1978) refer to as double-loop learning? Double-loop learning is defined as "those sorts of organizational inquiry that resolve incompatible organizational norms by setting new priorities and weightings of norms, or by restructuring the norms themselves with associated strategies and assumptions." Double-loop learning involves surfacing and challenging deep-rooted assumptions and norms of an organization that have previously been inaccessible.

The learning laboratory can help advance double-loop learning by providing a unique forum in which operating norms and assumptions can be questioned in a non-threatening way, by experimenting with the computer simulator. For example, in a company that professedly emphasized pursuing high quality standards, the behavior in the simulator trials showed that controlling expenses dominated people's actions. One manager remarked that while playing the game "I kept telling myself, 'don't add to staff, don't add to staff,' even though there is no one telling me not to *and* knowing that I really need to!" In many cases, when people discovered they had extra capacity, they chose to cut staff or push for more production to reduce expenses, instead of pushing for quality.

When given an office setting where quality was twice as high as in all previous scenarios, the vast majority decided to cut staff. In the short term, they reaped the benefit of lower expense costs, but those savings were dwarfed by the heavier losses incurred due to poor quality. In the insurance company claims learning lab, when asked why they had decided to cut staff, most reply that they felt the office was like a country club. They point to the number of claims in "backlog per adjuster" and number of "claims settled per

adjuster" and state that the workload was incredibly light. They express matter-of-factly that the right number should be twice what they were. When asked what made those the right numbers, they invariably pause, and reply sheepishly with the only answer that they can, "Because it's always been that way." In that instant, they realize how an unquestioned assumption had driven their decision making and may have contributed to poor performance.

The use of management flight simulators in learning laboratories advances double loop learning by providing people with a framework for clarifying their mental models. Through the process of cycling through many rounds of decision making and rethinking their mental models, managers gain insights into their domain-specific issues. Although we have only anecdotal evidence linking "better management thinking" to better total cost or profit numbers, learning laboratories have great potential for helping managers reassess the way they think about their business. In the following section, we begin to address the measurement of this potential.

4. Measurement of Flight Simulator Learning

Although the learning laboratories appear to have helped managers be more aware of their own mental models, further evidence is needed to determine the effectiveness of management flight simulators in improving actual decision making. A three-day system dynamics seminar for managers, held at a major computer manufacturer, has provided the opportunity to pilot test a number of instruments that were designed to gather experimental data about learning transfer after exposure to a system dynamics seminar and learning laboratory. The goal of the system dynamics seminar is for the managers to learn the concepts and then transfer them back to the work place.

4.1. Learning Objectives

The managers used the People Express Airline Management Flight simulator (see Graham et al. in this journal for a description of this simulator) during the seminar. They also read the case materials, and were taught to use causal loop diagrams to conceptualize about the feedback structure of People Express organization and its competitive environment. We expected to observe three levels of learning: understanding of the People Express case issues, understanding of the underlying feedback structure, and transferring insights to another domain.

Attaining the first level of learning means that managers have identified the basic organizational issues that led to the failure of People Express. They may explain, for example, that the low price offered by People Express created a huge demand for flights

that could not be fulfilled with their existing service capacity. Their failure to expand service capacity rapidly led to poor service quality that eventually resulted in the demise of People Express.

The second level of learning requires managers to explain at a deeper level, linking the observed behavior to the underlying feedback structure of People Express. They should be able to articulate short-term versus long-term tradeoffs in the system and recognize the link between system structure and dynamic behavior.

At the third level of learning, managers should be able to transfer the systems insights gained from studying People Express to other cases. For example, they should be able to recognize the same "attractiveness principle" (described below) that operated in the People Express case when it surfaces in another case with a different cover story (i.e., not about an airline). For this article, we focus on this third level of learning and the role of the management flight simulator.

4.2. Effectiveness

The issue of effectiveness of the seminars and learning laboratory arises because we do not understand *which* thinking processes change nor *how* they are changed by the training intervention. Unlike basic skills training, such as typing or computer repair, where success is measured by the student being able to perform the task, learning of system dynamics principles is not as easily evaluated. System dynamics training encompasses multiple levels of knowledge acquisition: from the basic skills level to higher levels of abstraction.

Measuring learning along certain dimensions is easier than along others. At a basic skills level, instruments can be used to measure a manager's ability to create a simple causal loop diagram. At a more abstract level, determining the manager's understanding of a complete feedback structure and his ability to transfer that knowledge to another case is difficult to measure. In the People Express management flight simulator, cumulative net income is a good indicator of *game* performance, but *management* performance is not easily quantified in this way. Management performance is multi-faceted and dynamic. Measurements must capture the skills developed from the basic to the abstract, as well as the transformation of the mental model as new skills are learned.

The research on simulation and games has generated controversy about the role and effectiveness of simulations in management education. Effectiveness of simulations is yet not proven. Early research on the transfer of case knowledge, in which no simulation or game is used, has shown that little transfer occurs when analogous problems are presented concurrently. Even less transfer may occur when problems are separated by a span of

months or years. This suggests that prior experience in solving problems may not be used by managers faced with new manifestations of analogous problems. Also, knowledge gained from experience may be misapplied because irrelevant associations may be formed between a term mentioned in a case problem and a term in an existing problem (Kardes, 1987).

Case analysis alone may not be sufficient for transfer of learning, but we suggest that a case analysis coupled with a game or simulator may improve the likelihood of transfer. Other evidence has suggested that participation in a management game provides increased learning, more favorable attitude, and higher levels of interest and motivation that are not achieved when working with a case alone (Raia, 1966). Further research will determine the degree to which a game approximates real-world policy making situations, if students learn the "right" things from the game experience, and whether game knowledge is efficiently transferred to the real world (Wolfe, 1976).

According to Wolfe, effectiveness research must deal with all the different situation variables that have an impact on the gaming application. These include 1) game design characteristics, such as single function versus functionally integrative, complexity, algorithm validity/face validity, random events; 2) administration characteristics, including starting position, team size, team selection, team accountability, duration, pacing, trial or practice runs, debriefing, within-course placement, learning objectives; 3) player and group characteristics, consisting of motivation, aptitude and achievement, attitude, cognitive style, participation, decision-making method, team structure; and 4) administrator characteristics, such as game experience and involvement, motivation, subject matter familiarity.

4.3. Instrument Design

The system dynamics seminar, the People Express simulator, and the instruments were designed to respond to many of Wolfe's criticisms of earlier effectiveness research. For example, the People Express game was tested with many managers and students to make the interface both simple to use and realistic in emulating a typical management information system. The set-up of the seminar allowed for random assignments of teams and debriefings at each transition. Surveys and questionnaires were used to gather data about player characteristics (age, education, and prior systems experience), measure attitude, and gather some data on cognitive style. The seminar and game administrators were specifically chosen for their substantial system dynamics teaching experience and their high-level of motivation.

The instruments used include surveys, questionnaires, strategy sheets, verbal protocols during game plays, game performance data, and causal loop diagrams. The

instruments have been designed to gather data on 1) prior experience with system dynamics; 2) case understanding (using a written examination which measured knowledge of the People Express domain); 3) performance in the People Express management flight simulator (for example, a comparison between groups of the average cumulative net income at the end of a game); 4) verbalization of systems concepts using protocol analysis while playing the game; 5) strategy sheets used while playing the game, which can demonstrate the use of the system dynamics framework when other measures may not; 6) a team's development of causal loop diagrams (representing their understanding of the feedback structure); 7) the transfer of systems insights about People Express to other cases; and 8) reflections on learning after participating in the entire seminar (Gould, 1989).

4.4. The Three Questionnaires

Although many instruments were used, we limit our discussion to three specific questionnaires designed to test for learning transfer. Figure 2 shows the sequence of the seminar and when the questionnaires (Q1, Q2, and Q3) were administered. The difficulty of transferring insights from the People Express case to other cases increases with each subsequent questionnaire.

The first questionnaire was designed to capture the manager's ability to recognize the most important issues faced by People Express after reading the case-study materials and participating in a discussion. The manager was then expected to transfer the lessons learned from the People Express case materials to another case. Examples of the questions used in the questionnaires are found below.

The second questionnaire was administered after Group A conceptualized about People Express (i.e., created the causal loop diagrams) and after Group B played the People Express game. Questionnaire 2 was designed to demonstrate the manager's ability to apply knowledge of the system dynamics framework to People Express as well as transfer the framework to another case. Each Group had a different experience before they answered questionnaire 2. This design allows us to compare the results based on each group's experience with the system dynamics framework. The conceptualizing process makes the framework an explicit part of the discussion. Playing with the management flight simulator does not include a discussion of how the flight simulator was designed using the system dynamics framework.

The third questionnaire was used after each group had conceptualized and played the game. This allows us to determine any effects from the sequence of experience using the system dynamics framework.

Group A	Introduction to Systems Thinking	Q1	Introduction to People Express	Conceptualize	Q2	Play game	Q3	Advanced Causal Loop Diagraming session
Group B	Introduction to Systems Thinking	Q1	Introduction to People Express	Play game	Q2	Conceptualize	Q3	Advanced Causal Loop Diagraming session

Figure 2. The two-group, pretest/posttest research design.

There are fourteen questions on the three questionnaires. There are four different types of questions as diagrammed in Figure 3. We would expect that the quadrant II and III questions would be easier for the manager to answer correctly than quadrant IV questions. Each question was designed to measure the transfer of the system dynamics framework from the original People Express case to the new cases.

	Same Cover Story	Different Cover Story
Same Feedback Structure	I	II
Different Feedback Structure	III	IV

Figure 3. Types of questions asked of the participants of the seminar.

It is expected that questions with the same cover story and feedback structure encourage transfer more than questions with a different cover story and different feedback structure. Questions with a similar cover story but a different feedback structure may be problematic due to misperception of the feedback structure because of the similarity of the cover story. When the cover story is different, but the feedback structure is the same, the managers may have trouble transferring their knowledge of the feedback structure because they are mislead by the new cover story. When the cover story and feedback structure are different from the original case, the new case may be perceived to be an entirely new problem with no relationship to the original case, and thereby may decrease the likelihood of transfer.

In addition to the four types of questions in Figure 3, the first question on each questionnaire was:

Why did People Express fail?

This question was used to measure any changes in response resulting from participation in the different phases of the seminar.

Each case example on the questionnaires was designed to reveal an underlying system dynamics feedback structure. The managers were taught about the "attractiveness principle" structure of People Express by participating in the discussions in the seminar and by playing the game. The "attractiveness principle" states that compensating changes in the components of attractiveness explain many past failures; frequently one aspect of a system is improved only to discover that other aspects have become worse. In the People Express case an increase in the availability of flights reduced the quality of service because of the increased demand and shortage of service capacity. Because organizational resources are limited, it is impossible to increase the quantity and quality of everything for everyone. It is only through the deliberate manipulation of counterbalancing effects that control can be gained over the changing system. The attractiveness of flying People Express lead to exponential growth in demand for seats. Many pressures develop to stop growth; some we can influence others we cannot. If pressures are alleviated where possible, growth will continue until it produces a further rise in the pressures that cannot be controlled (Mass. 1974).

If people actually learned the attractiveness principle, we would expect them to apply that knowledge to the other cases on the questionnaires. Examples of two questions used in the questionnaires follow.

Each question on the questionnaires was designed to be an example of a short case. The following question is an example of a question from quadrant II. It has a similar feedback structure to People Express ("attractiveness principle"), but a different cover story (i.e., not about an airline).

> Commuting by car to downtown Boston from the suburbs takes much longer now than it did a decade ago. The state is now planning to construct a new expanded highway system to increase access to the downtown area. What effects will this have on the commute time? Is this case related in any way to the People Express Case? Explain.

An adequate response to this question would show an understanding of the "attractiveness principle," and would lead the manager to suggest that overcrowding will be reduced in the short-term, but in the long-run the highway will again be crowded because of the increased attractiveness of commuting on the new highway.

The question below is an example of a question from quadrant IV. It has a similar cover story to People Express, but a different feedback structure.

> Northeast Airlines has a small fleet that serves passengers travelling to small airports throughout the New England, New York, Pennsylvania and New Jersey area. Northeast is known as a high quality airline, but has grown very slowly during the past decade. The Vice President just left the company after a dispute with the CEO. The VP had tried to persuade the CEO to expand into an airport that was to be available soon in southern New Hampshire. This region was not served by Northeast. The CEO resisted. An Air Force base in southern New Hampshire will be decommissioned soon, and will be converted for civilian use. Commercial airlines will be able to lease terminal space. The landing field will accommodate planes as large as a DC8. Northeast has enough lead-time for planning their entrance into the market. Preliminary market research shows that Northeast has a good reputation within the new market. Northeast will need to add staff and purchase more planes for a total expansion of about 50%. The CEO thinks this expansion is too big for the company to absorb. He believes Northeast's high quality over the past decade is a result of his cautious, slow growth strategy. Should the CEO have listened to the VP's advice? Explain. Is this case related in any way to the People Express case? Explain.

Unlike People Express, where the "attractiveness principle" dominates, the former VP of Northeast airlines may have suggested the correct policy. By expanding capacity before the increase in expected sales of seats, the VP may have expected that attractiveness in a new market with the proper advanced planning would lead to a successful business and lessen the immediate effect of the "attractiveness principle."

4.5. Preliminary Results

Pilot testing the instruments described above has helped identify areas of weakness and opportunities for improvement. The method of repeating the same question on each questionnaire (i.e., Why did People Express fail?), for example, led the managers to believe that a change in their response was expected on each questionnaire and could therefore bias the results. The wording of the simple case questions, which were designed to measure transfer, could be misleading. For example, not all managers would agree that the People Express and the Northeast Airlines cases represent examples of organizations with a similar cover story, but a different underlying structure. The instruments need further refinements to address these issues.

Despite these problems with the instruments, an initial review of the questionnaires shows that some managers are able to use the system dynamics feedback terminology and demonstrate evidence of system dynamics thinking. For example, the question about the Boston highway expansion yielded this response from one manager:

> In a short-time more cars will fill the highway with a minimal decrease in commute time. It may even take longer as better highways increase housing development.

As anticipated, the response to the Northeast Airlines question showed little transfer.

In addition to the problems with the instrument design itself, the corporate setting creates many measurement control problems. The shortage of computers requires that managers work in teams while playing the game. The team discussions may improve performance for some managers and not for others. Managers frequently leave the seminar to respond to inquiries. Therefore, there is no assurance that all managers receive the same treatment. These issues further compromise our ability to separate the effects of playing the game versus conceptualizing. Although evidence of transfer appears on the questionnaires, instrument redesign and a controlled research setting may further our understanding of the effects of the system dynamics framework on management performance.

The next section of this paper uses management flight simulators with managers and students in a laboratory setting, which improves the experimental control. The simulators and debriefings are the only intervention and therefore allow us to determine more precisely the effect of the simulator on the transfer of learning across different settings.

5. Transfer of insights

Enriching professionals' mental models as far as dynamic interactions are concerned is not an end in itself. On the contrary, we can only argue that increasing people's ability to grasp dynamic complexity is desirable if it serves some concrete purpose. The purpose is, of course, increased competency and improved organizational decision making. Both benefits assume that people can transfer insights achieved in a laboratory situation. There are two dimensions to this transfer, transfer from workshop to work place and transfer from one workshop environment to another. The first transfer dimension depends on, but is not limited to, the second. In this section, we describe an experiment that investigates how students and professionals transfer differently from one Management Flight Simulator to another.

MBA students and young professionals interacted with real estate and oil tanker simulators in an experimental study. These two markets were chosen for several reasons. First, time lags create substantial delays between desired and actual stocks of capacity in both cases. Second, both markets exhibit a short-time decision focus. Third, learning is problematic and market instabilities may persist for long periods of time (Hernandez, 1991).

5.1. Experimental and Real Markets

Two computerized decision-making games were designed with cover stories depicting them as a world market for oil tanker transportation and office real estate. The feedback structures underlying both games (shown in Figure 4) are identical. The dynamic behavior arises from the fact that investors invest when expected operating profits are high and, conversely, do not invest when low profits are expected. Instabilities in these markets are accentuated by participants not accounting sufficiently for assets under construction (Hernandez, 1991; Sterman, 1989). In both markets demand is insensitive to price fluctuations whereas supply is very price sensitive. This causes instability and over-investment.

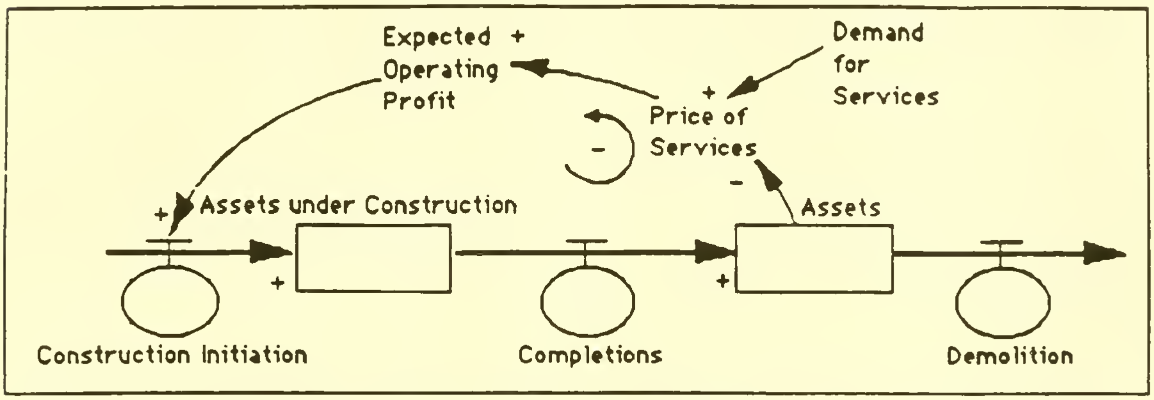

Figure 4. The main feedback structure of the market. + signs mean that variables change in the same direction. - signs mean that variables change in opposite direction. The negative "loop" indicates that the feedback loop is goalseeking (seeking balance between demand and supply of services).

The simulated markets were identical with the exception of variable names and parameter values. One such difference is the higher substitutability in the real estate market, which makes this market more stable than the tanker market. Another difference is the average completion time for "Assets under Construction", which is longer in the tanker market. As a consequence, the shipping market exhibits more pronounced fluctuations. Though the construction cycle is far from a well-tuned pendulum clock, Figures 5 and 6 show that the element of industry cyclicality has persisted over the years. The results generated in the flight simulators were very similar to these historical time series.

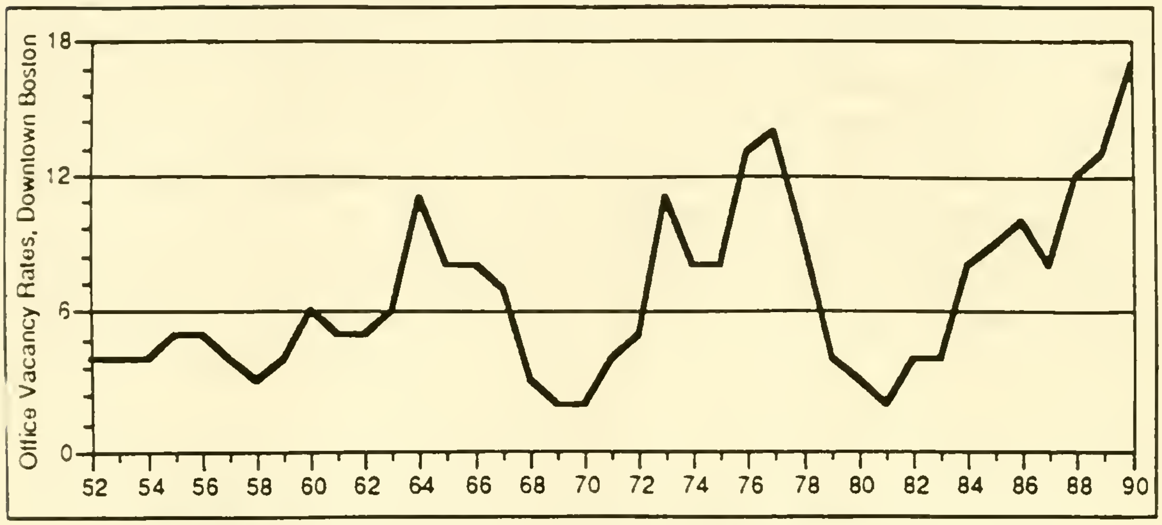

Figure 5. Percent office vacancy rates, Boston. Source: Boston Redevelopment Authority (1989)

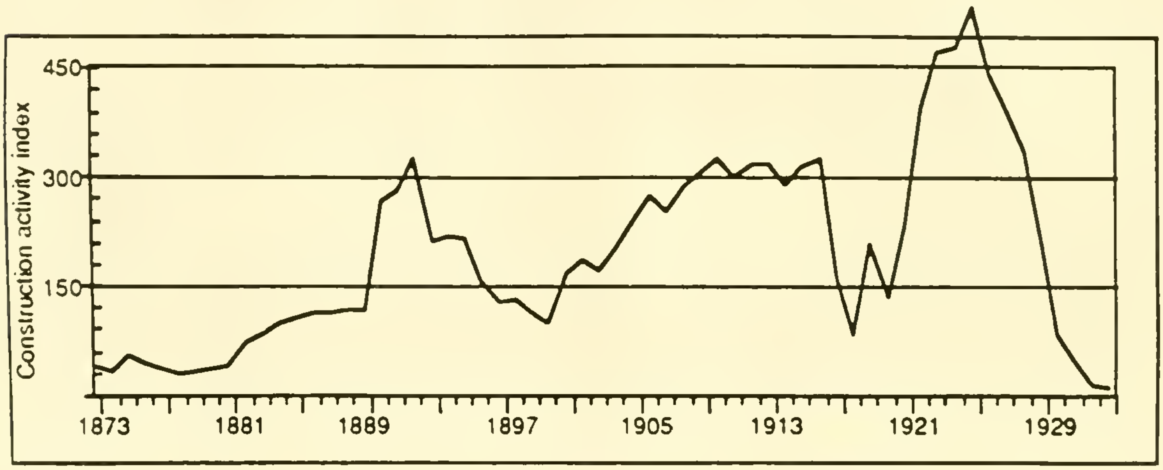

Figure 6. Construction activity index, Chicago. Source: Hoyt (1933).

The market for oil tanker transportation exhibits dynamics that are similar to those of the real estate industry, but instabilities are more dramatic. Prices show periods of calm when there exists slack capacity. However, when capacity utilization creeps above equilibrium, prices become highly erratic. In fact, when capacity utilization is low, the market is equilibrated by "mothballing" tankers, but when capacity utilization is high, transportation rates fluctuate to clear the markets.

5.2. Procedure

In the study reported here, the players were mostly current or future MBA's. Seventeen were students at MIT's Sloan School of Management. The students participated

individually and were paid according to their performance. The other group consisted of 32 professionals from major corporations. Most had two to ten years' work experience in addition to an MBA degree. Several had long experience in one of the two markets. They played in *teams of two*, and the results of the 16 teams are included here.

The students were given monetary incentives, whereas the professionals took part in day-long seminars where the game-playing was an early, yet integral part of reflective exercises that culminated in a discussion session. Since these discussions took place after game-playing, the game-playing procedure itself was identical and comparable across both groups.

All participants first read a two-page newspaper article about market conditions portrayed in the first game. In addition, all subjects read a briefing book about the simulator before interacting with the computer. Then the game-playing started. Subjects were randomly assigned to the real estate or the oil tanker condition. Players were instructed to start again in the event of bankruptcy. If they did not go bankrupt, they continued until the game stopped. All full trials lasted 40 periods (the players were not informed in advance about the length of a trial). Two complete trials were completed within the first market condition. After a break, all players were given a two-page article about the other market, i.e., the transfer market, and asked to participate in that market until they had completed at least one full trial.

All participants were asked to maximize profits in each game. The student players were rewarded for average performance in all games they played ($4 per hour). In addition, they were given a performance bonus that amounted to $110 for the participant with the highest average profit. Instead of monetary rewards, the professionals had to announce their results publicly to their peers after the sessions, and that was apparently a strong incentive to perform well.

5.3. Results

Game performance can be a good indicator of how well players understand a cyclical market. Research in a similar setting has shown a positive relationship between performance and understanding of underlying feedback structure (Bakken, 1990). In Figure 7, we see that students go bankrupt twice as frequently as professionals before they finish their first full, 40-decision, trial. Analysis of other potential differences, such as real world experience in the simulated market, did not make any difference in game-playing strategy or outcome. Additionally, results showed no difference between subjects who started with Real Estate and continued with Oil Tankers and those who played the games in the reversed order.

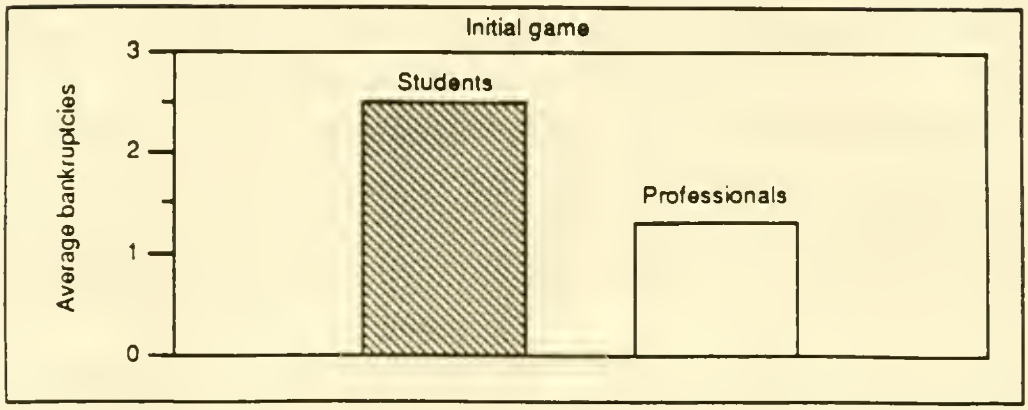

Figure 7. Students go bankrupt more often.

Figure 8 shows that average performance in early trials is similar for both groups, and that both groups have transferred insights from the first to the second simulator. Transfer performance among the students, however, is markedly better than the professionals. The higher number of bankruptcies by the student group indicates that a different strategy was employed. In general, players who go bankrupt make more decisions than people who succeed in finishing a full round without going bankrupt. A high number of bankruptcies indicates that a player has an exploratory attitude and a willingness to make risky decisions. Bankruptcies impede performance initially, but provide valuable decision-making experience. In the long run, this exploratory decision-making style helps develop a deeper understanding that proves beneficial in the transfer game.

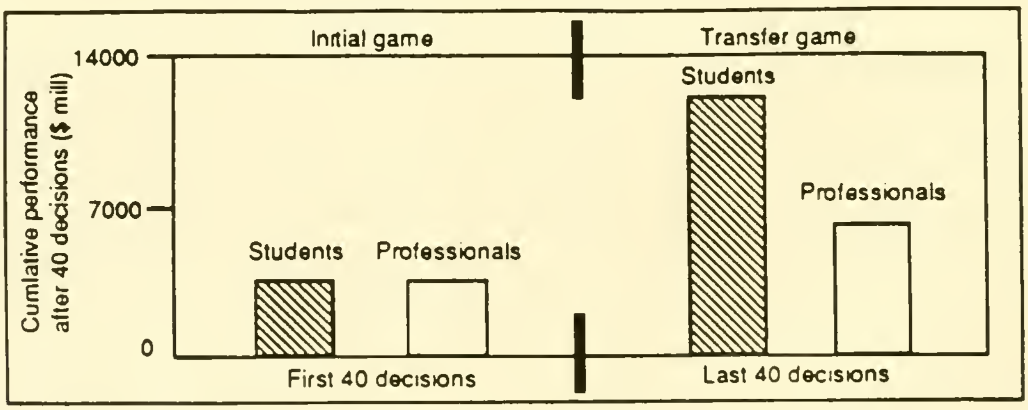

Figure 8. Students do as well as professionals initially, and transfer better.

5.4. Why Are Professionals Less Exploratory?

What explains the apparent differences in decision-making strategy between the students and the professionals? One reason could be that students played individually while professionals in teams of two. Since aversion to risk is higher in team decision making, the consensus needed for team decision making reduces risk to the most risk-averse of the team players. In addition, our experimental results indicate that although every effort was made to ensure a non-threatening environment, professionals still feel restricted. As we mentioned in the previous section, participants tend to use known (real world) strategies to anchor their decisions. The same embedded routines that hinder exploration in the real world can also prevent exploration and learning in the simulated environment.

Another possible explanation for the difference in exploratory attitude is that the simulated environment poses more threat to professionals than to students. A hidden underlying assumption appears to be: "I know a lot about this market, so I just do in the game what I do in real life." Though fairly safe, such an attitude is counterproductive to learning since it fails to benefit from the experimental approach facilitated by the laboratory setting. Students' game playing strategy, on the other hand, seems to be more playful. They appear to use the information available in the game as a springboard for investigation into the causal dynamics. Not surprisingly, students take the exercise more as a learning experience. In the very short run they suffer and go bankrupt; in the long run they improve their performance.

The simulated environment is threatening enough for professional participants to limit their decision repertoire and impede learning as well as the ability to transfer. Threat has the general effect of narrowing options considered. Thus, without embedding the management flight simulator in a context that makes the simulator less threatening (such as a learning lab), the results indicate that real world learning disabilities will be reproduced in a simulated world.

5.5. Poor transfer of learning; what can and should be done?

The professionals spent several hours in directed discussions after the experimental game-playing sessions that helped remedy some of the transfer shortcomings of the simulator. First, experiences were shared among the teams, enabling those who had followed conservative strategies to benefit from those who had pursued more adventuresome decisions. In a very compressed time interval, the discussion revealed that markets that seem to be different, in fact, share many commonalties. Participants could

quickly perform mental what-if analyses due to the intense common experience of the simulator.

The simulator experience appeared to facilitate the sharing of real-life experiences. For example, professionals in one learning laboratory with real estate experience in the northeast region questioned the fact that asset values in the simulator could swing as much as 40% between peaks and troughs. During the last ten years, they argued, asset values had never declined, and they knew of no other instance where values had dropped by more than 10%. Interestingly, there were people who had lived through depressed market conditions in the southern states in the same session. They questioned the validity of the simulated asset cycles because they had themselves experienced asset value reductions of 50% over a period of a few years! The discussion then turned towards the preconditions for such a fall and what it meant for the northeast region.

From an organizational learning viewpoint, it is interesting that the colleague with exposure to the market with the 50% drop had not previously shared his experiences. Presumably, he had, until the simulated experience, thought that the southern market was so unique that it was irrelevant to current colleagues. The management flight simulator gave added credibility to viewing these markets as cyclical. Furthermore, the learning laboratory provided a framework where both commonalties and differences between cyclical markets could be integrated. This was done by the facilitator when the feedback structure depicted in Figure 4 was shared among the participants.

One real estate development company used the learning laboratory to instigate changes in its incentive structure. These changes were predicated on expected future difficulties induced by the cyclical nature of their business. Until then, it had been hard to convince junior partners that focusing exclusively on developing real estate was problematic. The lesson taken from professionals with three to ten years of experience was, "An investment missed today is millions lost in unearned capital gain tomorrow." The robustness of such a lesson is greatly reduced if the relevant time horizon is 30-40 years rather than 5 years. The cyclical experiences and emotionally laden bankruptcies improved the receptiveness for organizational change.

5.6. Theoretical interpretation

Work on simple algebra word problems has shown that people have difficulties applying problem solving insights from one domain to another. Risking oversimplification, one can say that research on transfer in this tradition has shown that though difficult to achieve, transfer of insights and strategies do occur. To transfer successfully, people need to possess a generic framework as well as experience in its use.

Frameworks are generic tools for a problem area, such as economics or statistics (Nisbett et al., 1986). But, a framework only helps transfer if its relevance is clear to a problem solver. In general, that is not the case, people do not see the relevance of general frameworks when decision situations arise. In our case, the students do not seem to grasp the relevance of the framework when they play the first trial of the first game. However, when they are exposed to later trials of the first game, they slowly realize the system structure. Bankruptcies increase decision exposures. When the second market is presented to them, the student players appear to understand the generality of underlying causal forces and act accordingly. Numerous studies in transfer of problem solving (Bassok and Holyoak, 1989; Gick and Holyoak, 1983; Gilden and Profitt, 1989; Newell and Simon, 1972; Simon and Hayes, 1982) show that people can indeed apply a framework but only after being prompted or after repeated exposures to a given problem structure.

There exists evidence that if people are exposed to a problem area under many different conditions they may develop their own frameworks if given time and if they see the relevance of doing so. As an example (Gentner and Toupin, 1986), most subjects are unable to recognize the relevance of a recently solved isomorphic problem when faced with a second one. However, when the subjects see a third isomorphic problem they "see a light" and realize that the two former problems shared structural characteristics and those are the same as the structure of the third problem.

6. Conclusion and future research

We have seen that management flight simulators can be integrated into learning laboratories to foster inquiry skills. Such skills are critical to a learning organization (Argyris and Schön, 1978; de Geus, 1988; Stata, 1989). People need richer mental models for domains where such mental models tend to be deficient. Deficiencies typically occur if decision making and feedback about the appropriateness of those decisions are remote in time and space. We have seen that by designing non-threatening learning laboratories, and by fostering experimentation in other ways, genuine reflection and experimentation may result. The effect of such experimentation is increased transfer, especially if accompanied by frameworks that can function as a vehicle for applying insights to new domains. The measurement of increased transfer and higher order thinking processes, is complicated. However, we have shown that such measures are possible.

The results reported here are not entirely conclusive. Further methodological refinements of the measurement instruments need to be made. In addition, more decision making, learning, and transfer of learning data need to be collected to substantiate and

validate the tentative conclusions presented in section 5. These validations are currently under way and will be made available from the first author in the coming months.

References

Argyris, C and Schön D (1978), *Organizational Learning: A Theory of Action Perspective,* Reading, MA: Addison-Wesley.

Bakken, B (1990), An interview with a thoughtful shipping investor. Working paper D-4167, System Dynamics Group, Sloan School, MIT.

Bassok, M and Holyoak K J (1989), Inter-domain transfer between isomorphic topics in algebra and physics. *Journal of Experimental Psychology; Learning, Memory, and Cognition* 15, 1, 153-166.

Boston Redevelopment Authority (1989), *The Boston Class A Office Market,* 363.

Brehmer, B (1989), Feedback delays and control in complex dynamic systems. In Milling and Zahn, eds., *Proceedings of the 1989 System Dynamics Conference.* Berlin: Springer Verlag.

de Geus, A (1988), Planning as learning. *Harvard Business Review* November-December.

Einhorn, H J and Hogarth, R M (1978), Confidence in judgment: persistence of the illusion of validity. *Psychological Review* 85.

Genter, D and Toupin, C (1986), Systematicity and surface similarity in the development of Analogy. *Cognitive Science* 10, No 3, 277-300.

Gick, M and Holyoak K J (1983), Schema induction and analogical transfer. *Cognitive Psychology* 15, 1-38.

Gilden, D L and Profitt (1989), Understanding collision dynamics. *Journal of Experimental Psychology; Learning, Memory, and Cognition* 15, 2, 372-383.

Gould, J M (1989), Tools for learning to manage a complex environment, working notes D-4051, System Dynamics Group, Sloan School, MIT.

Hernandez, K L (1991), Learning in real estate: the role of the development system in creating oversupply, Unpublished M.S. Thesis, Sloan School, MIT.

Hoyt, H (1933), *One Hundred Years of Land Values in Chicago,* Chicago: University of Chicago Press.

Kahnemann, D and Tversky, A (1974), Judgment under uncertainty. *Science* 185.

Kardes, F (1987), The case method of instruction and managerial decision making, Working paper, Sloan School, MIT.

Kim, D (1989), Learning laboratories: designing reflective learning environments, In Milling and Zahn, eds., *Proceedings of the 1989 System Dynamics Conference.* Berlin: Springer Verlag.

Mass, N (1974), *Readings in Urban Dynamics.* Cambridge, MA: Wright-Allen Press.

Newell, A and Simon H (1972), *Human Problem Solving,* Englewood Cliffs, NJ: Prentice-Hall.

Nisbett, R; Fong, G; Lehman, D and Cheng, P (1987), Teaching reasoning. *Science* 238, 625-631.

Papert, S (1981), *Mindstorms.* New York: Basic Books.

Raia, A P (1966), A study of the educational value of games. *Journal of Business* 39(3), 339-352.

Schön, D A (1983), *The Reflective Practitioner: How Professionals Think in Action.* New York: Basic Books.

Senge, P M (1990), *The Fifth Discipline.* New York: Doubleday.

Simon, H (1956), *Administrative Behavior.* New York: Wiley and Sons.

Simon, H (1978), Rationality as process and product of thought. *American Economic Review* 68.

Simon, H and Hayes J R (1974), Understanding written problem instructions. In L W Gregg, ed., *Knowledge and Cognition.* Potomac, MD: Lawrence Erlbaum.

Stata, R (1989), Organizational learning the key to management innovation. *Sloan Management Review,* October.

Sterman, J (1987), Testing behavioral simulation models with direct experiment. *Management Science* 33, 1572-1592.

Sterman, J (1989), Misperceptions of feedback in dynamic decision making. *Organizational Behavior and Human Decision Processes* 43, 301-335.

Vennix, J (1990), *Mental Models and Computer Models*, Ph. D. Dissertation. University of Nijmeigen, Center for Cognition and Information Research.

Weick, K E (1977), Enactment processes in organizations. In Staw and Salanchick, eds., *New Directions in Organizational Behavior*, Chicago: St. Clair Press.

Wolfe, J (1976), The effects and effectiveness of simulations in business policy teaching applications. *The Academy of Management Review* (1) 2, April.